

Uncertainty Estimates in the Heston Model via Fisher Information

Oliver Pfante and Nils Bertschinger

Systemic Risk Group, Frankfurt Institute for Advanced Studies,
Routh-Mofang-Straße 1, 60483 Frankfurt am Main

October 19, 2016

Abstract

We address the information content of European option prices about volatility in terms of the Fisher information matrix. We assume that observed option prices are centred on the theoretical price provided by Heston's model disturbed by additive Gaussian noise. We fit the likelihood function on the components of the VIX, i.e., near- and next-term put and call options on the S&P 500 with more than 23 days and less than 37 days to expiration and non-vanishing bid, and compute their Fisher information matrices from the Greeks in the Heston model. We find that option prices allow reliable estimates of volatility with negligible uncertainty as long as volatility is large enough. Interestingly, if volatility drops below a critical value, inferences from option prices become impossible because Vega, the derivative of a European option w.r.t. volatility, nearly vanishes.

Index terms— Fisher information; Stochastic Volatility; Greeks

1 Introduction

Volatility of stock processes is a highly volatile time-process itself. This insight led to the introduction of volatility indices like the VIX (1993) and its off-springs, based on the work [8, 9], which make volatility a trademark in its own right subject to similar stochastic movements as stock prices. According to this view volatility seems to be responsible for several statistical properties of observed stock price processes. In particular, volatility clustering, i.e., large fluctuations are commonly followed by other large fluctuations and similarly for small changes [7]. Another feature is that, in clear contrast with price changes which show negligible autocorrelations, volatility autocorrelation is still significant for time lags longer than one

year [29, 28, 7, 25, 15, 24]. Additionally, there exists the so-called leverage effect, i.e., much shorter (few weeks) negative cross-correlation between current price change and future volatility [6, 7, 4, 5].

In stochastic volatility models, volatility is considered as a hidden process which can only be observed indirectly via its effect on stock price dynamics. Thus, in practice it has to be inferred from market data. In a previous paper [31], we have shown that daily stock returns provide only very limited information about volatility. Thus, there are intrinsic limits on how precisely volatility can be recovered from market data. Here, we consider a related question when recovering volatility from option price data. To our knowledge, there are no estimates of the uncertainty associated with volatility when inferred from option prices.

We tackle this question in terms of Fisher information which quantifies the uncertainty of a maximum likelihood estimate by the curvature of the likelihood function around its maximum. A shallow maximum would have low information as the parameters are only weakly determined; while a sharply peaked maximum would have high information. Further, the famous Cramer-Rao bound links Fisher information to the minimum variance of an unbiased estimator, thus providing fundamental limits on the reliability of parameter estimation.

In the present context, among the vast family of stochastic volatility models, we focus on Heston's model. It successfully models statistical properties of stock returns [13, 7, 16, 26] and the implied volatility surface [18, 19, 23, 32, 21, 22, 27]. We refer also to the introduction of [17] for a discussion of the empirical properties of Heston's model. There, the authors derive an analytical expression for the characteristic function of an extension of the Heston model in terms of fractional Brownian motions. Most important, along with Heston's original paper [20] came also an analytical expression of the European option prices in terms of characteristic functions and many algorithms have been proposed for their computation – see also Rouah's book [33] and the references therein for a thorough discussion of these algorithms.

We illustrate our results on an option portfolio which underlies the VIX index and compute the absolute and relative uncertainties of S&P 500 volatility over the last decade. Overall, we find that in contrast to stock returns alone [31], option prices provide substantial information about volatility making inferred volatility a precise estimator. Interestingly, very small volatilities are most difficult to estimate as Vega almost vanishes below a critical value. This in turn leads to huge relative errors in small inferred volatilities. The VIX, being a variance swap on the average volatility over 30 days, is much more stable with a relative uncertainty never exceeding 3% over the considered data set.

The paper is structured as follows: Sec. (2) introduces Heston's model and associated pricing formulas. Further, we explain how the fractional fast Fourier transform allows an efficient computation of the Heston Greeks. In Sec. (3) we state the formal definition and important properties of Fisher information. Sec. (4) illustrates how the Fisher information varies w.r.t. strike and maturity in the Heston model. Further, we discuss the role of Vega on the reliability of volatility estimation. Finally, in Sec. (5) we compute the Fisher information for the volatility of the S&P 500 index as well as the VIX index.

2 Pricing options in Heston's Model

2.1 Heston's Stochastic Volatility

Introducing Heston's model for pricing options we follow [33]. The Heston model assumes that the underlying stock price, S_t , follows a Black-Scholes-type stochastic process, but with a stochastic variance v_t that follows a Cox, Ingersoll, and Ross process. Hence, the Heston model is represented by the bivariate system of stochastic differential equations (SDEs)

$$\begin{aligned} dS_t &= \mu S_t dt + \sqrt{v_t} S_t dW_{1,t} \\ dv_t &= \kappa(\theta - v_t)dt + \sigma\sqrt{v_t}dW_{2,t} \end{aligned} \quad (2.1)$$

with the instantaneous correlation $dW_{1,t}dW_{2,t} = \rho dt$ of the two Brownian motions. The parameters of the model are

drift of the stock	μ
relaxation parameter	$\kappa > 0$
long-term mean of the variance	$\theta > 0$
leverage-effect parameter	$\rho \in [0, 1]$
volatility of the variance	$\sigma > 0$.

Furthermore, the price at time t of a zero-coupon bond paying 1\$ at maturity $t + \tau$ is

$$P(t, t + \tau) = e^{-\tau r}$$

with constant interest rate r . Neglecting volatility risk premium, change of measure yields the log-price $x_t = \log S_t$ and variance v_t

$$\begin{aligned} dx_t &= \left(r - \frac{1}{2}v_t \right) dt + \sqrt{v_t}d\tilde{W}_{1,t} \\ dv_t &= \kappa(\theta - v_t)dt + \sigma\sqrt{v_t}d\tilde{W}_{2,t} \end{aligned}$$

w.r.t. to the risk neutral measure \mathbb{Q} , see [33] or Heston's original work [20] for a detailed derivation. If we include a continuous dividend yield q , the time drift of the log-price becomes $r - q - 1/2v_t$.

2.2 European options

We present the price for a European call and put option in Heston's model in the formulation of Carr and Madan [10]. Henceforth, we abbreviate the log-price $x_t = \log S_t$ at time t simply by x and similarly the variance v_t at time t by v . We only consider the characteristic function f of the cumulative distribution $P^{\mathbb{Q}}(X_\tau > \log K)$ w.r.t. the risk-neutral probability, that is

$$P^{\mathbb{Q}}(X_\tau > k) = \frac{1}{2} + \frac{1}{\pi} \int_0^\infty \Re \left[\frac{e^{-i\phi k} f(\phi, x, v, \tau)}{i\phi} \right] d\phi.$$

with

$$f(\phi, x, v, \tau) = e^{(C(\phi, \tau) + D(\phi, \tau)v + i\phi x)}$$

and the logarithmic strike $k = \log K$. The little Heston Trap formulation [3] yields

$$\begin{aligned} C(\phi, \tau) &= i(r - q)\phi\tau + \frac{\kappa\theta}{\sigma^2} \left[(Q - d)\tau - 2 \log \left(\frac{1 - ce^{-d\tau}}{1 - c} \right) \right] \\ D(\phi, \tau) &= \frac{Q - d}{\sigma^2} \left(\frac{1 - e^{-d\tau}}{1 - ce^{-d\tau}} \right) \end{aligned}$$

with

$$\begin{aligned} c &= \frac{Q - d}{Q + d} \\ d &= \sqrt{Q^2 + \sigma^2(i\phi + \phi^2)} \\ Q &= \kappa - i\rho\sigma\phi. \end{aligned}$$

We introduce

$$E(\epsilon, x, v, r, \kappa, \theta, \rho, \sigma, k, \tau) = \frac{e^{-\epsilon\alpha k}}{\pi} \int_0^\infty \Re \left[e^{-ik\phi} \hat{e}(\epsilon, \phi, x, v, \tau) \right] d\phi \quad (2.2)$$

with

$$\hat{e}(\epsilon, \phi, x, v, \tau) = \frac{e^{-r\tau} f(\phi - i(\epsilon\alpha + 1), x, v)}{(\epsilon\alpha)^2 + \epsilon\alpha - \phi^2 + i\phi(2\epsilon\alpha + 1)}.$$

where $\alpha > 0$ is a positive damping factor which can be chosen according to an optimization scheme outlined in [10] and $\epsilon \in \{1, -1\}$. We obtain the European call $C(x, v, r, \kappa, \theta, \rho, \sigma, k, \tau)$ and put $P(x, v, r, \kappa, \theta, \rho, \sigma, k, \tau)$ price, respectively, via

$$\begin{aligned} C(x, v, r, \kappa, \theta, \rho, \sigma, k, \tau) &= E(1, x, v, r, \kappa, \theta, \rho, \sigma, k, \tau) \\ P(x, v, r, \kappa, \theta, \rho, \sigma, k, \tau) &= E(-1, x, v, r, \kappa, \theta, \rho, \sigma, k, \tau). \end{aligned} \quad (2.3)$$

We refer to [10] or the third chapter in Rouah's book [33] for a derivation of the formulae Eq. (2.3). Furthermore, we have put-call parity, see [33],

$$P(x, v, r, \kappa, \theta, \rho, \sigma, k, \tau) = C(x, v, r, \kappa, \theta, \rho, \sigma, k, \tau) + e^k e^{-r\tau} - e^x e^{-q\tau}. \quad (2.4)$$

The Heston Greeks in terms of the Carr-Madan formulation read

$$\begin{aligned} \frac{\partial}{\partial \gamma} E(\epsilon, x, v, r, \kappa, \theta, \rho, \sigma, k, \tau) \\ = \frac{e^{-\epsilon \alpha k}}{\pi} \int_0^\infty \Re \left[e^{-ik\phi} \mathfrak{f}_\gamma(\phi - i(\epsilon \alpha + 1), x, v, \tau) \hat{e}(\epsilon, \phi, x, v, \tau) \right] d\phi \end{aligned} \quad (2.5)$$

where \mathfrak{f}_γ are different functions for $\gamma \in \{\sigma_0, \kappa, \sqrt{\theta}, \rho, \sigma\}$ and $\sigma_0 = \sqrt{v}$ denotes the volatility.

2.3 Fractional Fourier Transform

Since the call and put price in Eq. (2.3) is expressed via a single Fourier integral we can apply a Fractional Fast Fourier Transform to achieve a simultaneous computation of the call and put prices Eq. (2.3) for various strikes. We follow the outline in [33]. We approximate the Fourier integral in Eq. (2.2) by Simpson's integration scheme over a truncated integration domain for ϕ , using N equidistant points

$$\phi_j = j\eta \quad \text{for } j = 0, \dots, N-1$$

where η is the increment. Simpson's rule approximates the integral Eq. (2.5) as

$$\begin{aligned} \frac{\partial}{\partial \gamma} E(\epsilon, x, v, r, \kappa, \theta, \rho, \sigma, k, \tau) \\ \approx \frac{e^{-\epsilon \alpha k} \eta}{\pi} \sum_{j=0}^{N-1} \Re \left[e^{i\phi_j k} \mathfrak{f}_\gamma(\phi_j - i(\epsilon \alpha + 1), x, v, \tau) \hat{e}(\epsilon, \phi_j, x, v, \tau) \right] w_j \end{aligned} \quad (2.6)$$

with the weight $w_0 = w_{N-1} = 1/3$, $w_j = 4/3$ iff j is odd and $w_j = 2/3$ otherwise. Since we are interested in strikes near the money the range for the log-strikes k needs to be centred on the log-price x . The strike range is, thus, discretized using N equidistant points

$$k_u = -b + u\lambda + x \quad \text{for } u = 0, \dots, N-1$$

where λ is the increment and $b = N\lambda/2$. This produces log-strikes over the range $[\log S - b, \log S + b - \lambda]$. For a log-strike on the grid k_u we can write

the sum Eq. (2.6) as

$$\begin{aligned}
& \frac{\partial}{\partial \gamma} E(k_u) \\
& \approx \frac{e^{-\epsilon \alpha k_u} \eta}{\pi} \sum_{j=0}^{N-1} \Re \left[e^{i \phi_j k_u} \mathfrak{f}_\gamma(\phi_j - i(\epsilon \alpha + 1), x, v, \tau) \hat{e}(\epsilon, \phi_j, x, v, \tau) \right] w_j \\
& = \frac{e^{-\epsilon \alpha k_u} \eta}{\pi} \sum_{j=0}^{N-1} \Re \left[e^{i j \eta (-b + u \lambda + x)} \mathfrak{f}_\gamma(\phi_j - i(\epsilon \alpha + 1), x, v, \tau) \hat{e}(\epsilon, \phi_j, x, v, \tau) \right] w_j \\
& = \frac{e^{-\epsilon \alpha k_u} \eta}{\pi} \sum_{j=0}^{N-1} \Re \left[e^{i \eta \lambda u j} e^{i(b-x)\phi_j} \mathfrak{f}_\gamma(\phi_j - i(\epsilon \alpha + 1), x, v, \tau) \hat{e}(\epsilon, \phi_j, x, v, \tau) \right] w_j
\end{aligned} \tag{2.7}$$

for $u = 1, \dots, N-1$. Applying a discrete Fast Fourier Transform (FFT) on Eq. (2.7) imposes the restriction

$$\lambda \eta = \frac{2\pi}{N} \tag{2.8}$$

on the choice of the increments λ and η which entails a trade off between the grid sizes. Hence, Chourdakis [12] introduced the Fractional Fast Fourier Transform (FRFT) to relax this important limitation of the discrete FFT. The term $2\pi/N$ in Eq. (2.8) is replaced by a general term β and Eq. (2.7) becomes

$$\hat{x}_u = E(k_u) \approx \frac{e^{-\epsilon \alpha k_u} \eta}{\pi} \sum_{j=0}^{N-1} \Re \left[e^{i \beta u j} x_j \right] \tag{2.9}$$

with

$$x_j = e^{i(b-x)\phi_j} \mathfrak{f}_\gamma(\phi_j - i(\epsilon \alpha + 1), x, v, \tau) \hat{e}(\epsilon, \phi_j, x, v, \tau) w_j \quad j = 1, \dots, N-1.$$

The relationship between the grid sizes λ and η becomes $\lambda \eta = \beta$. Thus, we can choose the grid size parameters η and λ freely. To implement the FRFT on the vector $\mathbf{x} = (x_0, \dots, x_{N-1})$ we first define vectors

$$\begin{aligned}
\mathbf{y} &= \left(\left\{ e^{-i \pi j^2 \beta} x_j \right\}_{j=0}^{N-1}, \{0\}_{j=0}^{N-1} \right) \\
\mathbf{z} &= \left(\left\{ e^{i \pi j^2 \beta} \right\}_{j=0}^{N-1}, \left\{ e^{i \pi (N-j)^2 \beta} \right\}_{j=0}^{N-1} \right).
\end{aligned}$$

Next, a discrete FFT of the vectors \mathbf{y} and \mathbf{z} yields $\hat{\mathbf{y}} = D(\mathbf{y})$ and $\hat{\mathbf{z}} = D(\mathbf{z})$ and we define the $2N$ dimensional vector

$$\hat{\mathbf{h}} = \hat{\mathbf{y}} \odot \hat{\mathbf{z}} = (y_j z_j)_{j=0}^{2N-1}.$$

where \odot denotes pointwise multiplication.

Now, we apply the inverse FFT on $\hat{\mathbf{h}}$ and take the pointwise product of the result with the vector

$$e = \left(\left\{ e^{-i\pi j^2 \beta} \right\}_{j=0}^{N-1}, \{0\}_{j=0}^{N-1} \right)$$

to obtain

$$\hat{\mathbf{x}} = e \odot D^{-1}(\hat{\mathbf{h}}) = e \odot D^{-1}(\hat{\mathbf{y}} \odot \hat{\mathbf{z}}) = e \odot D^{-1}(D(\mathbf{y}) \odot D(\mathbf{z})) .$$

If we truncate the last N elements we obtain the desired vector \mathbf{x} of Eq. (2.9) with length N . Hence, the FRFT maps a vector of length N onto another vector of length N , even though it uses intermediate vectors of length $2N$.

3 Fisher Information

3.1 Introduction

The outline on the Fisher information and the Cramér-Rao inequality follows [14]. We begin with a few definitions. Let $\{f(x; \Theta)\}$, $\Theta = (\theta_1, \dots, \theta_m) \in \mathcal{P} \subset \mathbb{R}^m$, denote an indexed family of densities $f : \mathcal{X} \rightarrow \mathbb{R}_{\geq 0}$, $\int f(x; \theta) dx = 1$ for all $\Theta \in \mathcal{P}$. Here \mathcal{P} is called the parameter set.

Definition 3.1. An *estimator* for Θ for sample size n is a function $T : \mathcal{X}^n \rightarrow \mathcal{P}$.

An estimator is meant to approximate the value of the parameter. It is therefore desirable to have some idea of the goodness of the approximation. We will call the difference $T - \Theta$ the error of the estimator. The error is a random variable.

Definition 3.2. The *bias* of an estimator $T(X_1, X_2, \dots, X_n)$ for the parameter Θ is the expected value of the error of the estimator, i.e., the bias is

$$\begin{aligned} \mathbb{E}_{\Theta} [T(x_1, x_2, \dots, x_n) - \Theta] \\ = \int (T(x_1, x_2, \dots, x_n) - \Theta) f(x_1; \Theta) \cdots f(x_n; \Theta) dx_1 \cdots dx_n . \end{aligned}$$

The estimator is said to be *unbiased* if the bias is zero for all $\Theta \in \mathcal{P}$ (i.e., the expected value of the estimator is equal to the parameter).

The bias is the expected value of the error, and the fact that it is zero does not guarantee that the error is low with high probability. We need

to look at some loss function of the error; the most commonly chosen loss function is the quadratic form

$$\begin{aligned} \Sigma(T(X_1, X_2, \dots, X_n)) = \\ \mathbb{E} \left[(T(X_1, X_2, \dots, X_n) - \Theta)(T(X_1, X_2, \dots, X_n) - \Theta)^T \right] \end{aligned} \quad (3.1)$$

which reduces to the covariance matrix for an unbiased estimator. Recall, covariance matrices are positive semi-definite and we write $A \leq B$ for two positive semi-definite $m \times m$ matrices if $x^T A x \leq x^T B x$ for all $x \in \mathbb{R}^m$.

Definition 3.3. An estimator $T_1(X_1, X_2, \dots, X_n)$ is said to *dominate* another estimator $T_2(X_1, X_2, \dots, X_n)$ if, for all Θ ,

$$\Sigma(T(X_1, X_2, \dots, X_n)) \leq \Sigma(T(X_1, X_2, \dots, X_n)).$$

The Cramér-Rao lower bound gives the minimum quadratic loss of the best unbiased estimator of Θ . First, we define the score gradient of the distribution $f(x; \Theta)$.

Definition 3.4. The *score gradient* V is a random variable defined by

$$V = \left(\frac{\partial}{\partial \theta_1} \log f(X; \Theta), \dots, \frac{\partial}{\partial \theta_m} \log f(X; \Theta) \right)$$

where $X \sim f(x; \Theta)$.

The expectation of every entry V_i of the score gradient is

$$\begin{aligned} \mathbb{E} V_i &= \int \left[\frac{\partial}{\partial \theta_i} \log f(x; \Theta) \right] f(x; \Theta) dx \\ &= \int \frac{\frac{\partial}{\partial \theta_i} f(x; \Theta)}{f(x; \Theta)} f(x; \Theta) dx \\ &= \int \frac{\partial}{\partial \theta_i} f(x; \Theta) dx \\ &= \frac{\partial}{\partial \theta_i} \underbrace{\int f(x; \Theta) dx}_{=1} \\ &= 0. \end{aligned}$$

Therefore, the covariance matrix of the score gradient V is $\mathbb{E} V V^T$. These entries have a special meaning.

Definition 3.5. The *Fisher information matrix* $J(\Theta)$ is the covariance matrix of the score gradient, i.e.,

$$J(\Theta)_{ij} = \mathbb{E} \left[\frac{\partial}{\partial \theta_i} \log f(x; \Theta) \frac{\partial}{\partial \theta_j} \log f(x; \Theta) \right]$$

The following properties of the Fisher information are crucial for the present paper. First, as covariance matrix, the Fisher information matrix is positive semi-definite. Second, information is additive: the information yielded by two independent experiments is the sum of the information from each experiment separately:

$$J_{X,Y}(\Theta) = J_X(\Theta) + J_Y(\Theta)$$

Third, the Fisher information depends of the parametrization of the problem: suppose Θ and Λ are m -vectors which parametrize the estimation problem, and suppose Θ is a continuously differentiable function of Λ , then

$$J(\Lambda) = D^T J(\Theta(\Lambda)) D \quad (3.2)$$

where the ij -th entry of the $m \times m$ Jacobian D is defined by

$$D_{ij} = \frac{\partial \theta_i}{\partial \lambda_j}$$

and D^T denotes the transpose of D . Finally, the significance of the Fisher information is shown in the following theorem.

Theorem 3.1. (*Cramér-Rao inequality*) *The covariance matrix of any unbiased estimator $T(X)$ of the parameter Θ is lower bounded by the reciprocal of the Fisher information:*

$$\Sigma(T(X)) \geq J(\Theta)^{-1}.$$

The Fisher information is therefore a measure for the amount of “information” about Θ that is present in the data. It gives a lower bound on the error in estimating Θ from the data.

3.2 Fisher Information in Pricing Options

The Fisher information matrix $J(\theta_{ij})_{ij \in \{\sigma_0, \kappa, \sqrt{\theta}, \rho, \sigma\}}$ of a single option with price e is a way of measuring the amount of information that the observable option price e carries about the unknown parameters $\sigma_0, \kappa, \sqrt{\theta}, \rho, \sigma$. Recall, σ_0 is the volatility, κ the relaxation parameter of the CIR process Eq. (2.1), θ the long-term mean of the variance, σ the volatility of the diffusion process Eq. (2.1), and ρ the leverage parameter, i.e., the instantaneous correlation between the two Brownian motions in Eq. (2.1).

While Heston’s option price $E(\epsilon, x, v, r, \kappa, \theta, \rho, \sigma, k, \tau)$ is a function of the stated parameters, the observed price e might deviate from its theoretical value. Following standard practice in fitting the volatility smile, we consider the mean squared loss between the actually observed option price e and its

theoretical value. In statistical terms, this corresponds to a noise model where the option price e is normally distributed around its theoretical mean value with variance \hat{v} . The probability function for e , which is also the likelihood function for the parameter vector $\Theta = (\sigma_0, \kappa, \sqrt{\theta}, \rho, \sigma)$, is a function $f(e; \Theta)$; it is the probability mass (or probability density) of the random option price e conditional on the value of Θ . The i, j -th entry of the Fisher information matrix is defined as

$$J(\Theta)_{ij} = \mathbb{E} \left[\frac{\partial}{\partial \theta_i} \log f(x; \Theta) \frac{\partial}{\partial \theta_j} \log f(x; \Theta) \right] \quad (3.3)$$

with $(\theta_1, \theta_2, \dots, \theta_5) = \Theta$. For a certain actual stock price S , strike K , and maturity τ we assume

$$f(e|\Theta) = \frac{1}{\sqrt{2\pi\hat{v}}} \exp \left(-\frac{(e - E(\epsilon, x, v, r, \kappa, \theta, \rho, \sigma, k, \tau))^2}{2\hat{v}} \right)$$

where $x = \log S$, $k = \log K$, and $v = \sigma_0^2$. We obtain the ij -th entry of the Fisher information matrix Eq. (3.3)

$$\begin{aligned} J(\Theta)_{ij} &= \mathbb{E} \left[\frac{\partial}{\partial \theta_i} \left(-\frac{1}{2} \log(2\hat{v}) + \frac{(e - E(x, \Theta, k, \tau))^2}{2\hat{v}} \right) \times \right. \\ &\quad \left. \frac{\partial}{\partial \theta_j} \left(-\frac{1}{2} \log(2\hat{v}) + \frac{(e - E(x, \Theta, k, \tau))^2}{2\hat{v}} \right) \right] \\ &= \mathbb{E} \left[\frac{e - E(x, \Theta, k, \tau)}{\hat{v}} \frac{\partial}{\partial \theta_i} E(x, \Theta, k, \tau) \times \right. \\ &\quad \left. \frac{e - E(x, \Theta, k, \tau)}{\hat{v}} \frac{\partial}{\partial \theta_j} E(x, \Theta, k, \tau) \right] \\ &= \frac{1}{\hat{v}^2} \mathbb{E} [(e - E(x, \Theta, k, \tau))^2] \frac{\partial}{\partial \theta_i} E(\epsilon, x, \Theta, k, \tau) \frac{\partial}{\partial \theta_j} E(\epsilon, x, \Theta, k, \tau) \\ &= \frac{1}{\hat{v}} \frac{\partial}{\partial \theta_i} E(\epsilon, x, \Theta, k, \tau) \frac{\partial}{\partial \theta_j} E(\epsilon, x, \Theta, k, \tau) \\ &= \frac{1}{\hat{v}} \nabla_{\Theta} E (\nabla_{\Theta} E)^T \end{aligned} \quad (3.4)$$

where

$$\nabla_{\Theta} E = \left(\frac{\partial}{\partial \theta_i} E(\epsilon, x, v, r, \kappa, \theta, \rho, \sigma, k, \tau) \right)_{i=1, \dots, 5}$$

is the gradient of the option price w.r.t. the parameters $\{\sigma_0, \kappa, \sqrt{\theta}, \rho, \sigma\}$. Furthermore, we consider deviations of observed prices of options with different strikes and maturities from their respective theoretical prices as independent. Additivity of Fisher information yields

$$J(\Theta) = \frac{1}{\hat{v}} \sum_{\tau \in \mathcal{T}, k \in \mathcal{K}} J_{\tau, k}(\Theta)$$

for simultaneously observed prices $\{E(\epsilon, x, v, r, \kappa, \theta, \rho, \sigma, k, \tau) : \tau \in \mathcal{T}, k \in \mathcal{K}\}$ of options with different maturities τ and log-strikes k .

4 Inferring hidden parameters from a single European option

According to Eq. (3.4), Fisher information matrix of a European option with Gaussian Noise centred on the theoretical price Eq. (2.2) provided by Heston's model is entirely determined by the first-order derivative of the option Price w.r.t. to the volatility σ_0 , and the parameters κ, θ, σ and ρ , respectively. We study these derivatives: first, their dependency of strikes and maturities; second, we have a closer look on Vega, i.e., the derivative of the option price w.r.t. to volatility σ_0 , observing a drop of its value for small volatilities.

4.1 Greek-Surfaces

The inverse of the Fisher information matrix $J(\Theta)$ in Eq. (3.4) provides a lower bound on the covariance matrix of an unbiased estimator $T(E)$ of the parameter $\Theta = (\theta_1, \dots, \theta_5) = (\sigma_0, \kappa, \sqrt{\theta}, \sigma, \rho)$ from a single option price E . Hence, the diagonal elements $J(\Theta)_{ii}^{-1}$, with $i = 1, 2, \dots, 5$, of the inverse $J(\Theta)^{-1}$ yield lower bounds of the variances of the estimates $T(E)_i$ of the parameters θ_i from a single observed option price E . The diagonal entries $J(\Theta)_{ii}^{-1}$ can be lower bounded by the respective Fisher information $J(\theta_i)^{-1}$. Note that $J(\theta_i)$ is the information obtained when estimating the parameter θ_i alone, i.e., considering all the other parameters as fixed. With this interpretation in mind the lemma below states that the variance of a joint estimator for all parameters Θ simultaneously is larger than estimating each parameter θ_i alone.

Lemma 4.1. *We have*

$$J^{-1}(\Theta)_{ii} \geq J(\theta_i)^{-1}$$

for $i = 1, \dots, 5$.

Proof. Let \mathbf{A} be an invertible $n \times n$ matrix and let denote \mathbf{A}_{ij} the ij -th block \mathbf{A} for $i, j \in \{1, 2\}$, i.e.,

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}.$$

Then, the inverse of \mathbf{A} can be expressed as, by the use of

$$\begin{aligned} \mathbf{C}_1 &= \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} \\ \mathbf{C}_2 &= \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}, \end{aligned}$$

as

$$\mathbf{A}^{-1} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{C}_1^{-1} & -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{C}_2^{-1} \\ -\mathbf{C}_2^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{C}_2^{-1} \end{bmatrix},$$

see equation 399 in [30]. Assume $\mathbf{A} = J(\Theta)$ and

$$\mathbf{A}_{11} = J(\Theta)_{11} = \frac{1}{\hat{v}} \left(\frac{\partial}{\partial \theta_1} E(x, k, \Theta, \tau) \right)^2.$$

Then

$$J(\Theta)_{11}^{-1} = (J(\Theta)_{11} - \mathbf{A}_{12} \mathbf{A}_{22} \mathbf{A}_{21})^{-1}$$

$J(\Theta)$ is positive semi-definite. This implies $\mathbf{A}_{12}^T = \mathbf{A}_{21}$, \mathbf{A}_{22} is positive semi-definite as well and therefore $\mathbf{A}_{12} \mathbf{A}_{22} \mathbf{A}_{21} \geq 0$. This yields the inequality for $i = 1$. Relabelling of the parameters yields the inequalities for the cases $i = 2, 3, 4, 5$ as well. \square

According to Eq. (3.4) we have

$$J^{-1}(\Theta)_{ii} \geq \hat{v} \left(\frac{\partial}{\partial \theta_i} E \right)^{-2} \quad \text{for } i = 1, \dots, 5.$$

Hence, the gradient $\nabla_{\Theta} E$ of an option price Eq. (2.2) does not only entirely determine $J(\Theta)$ but its components also provide first estimates on the uncertainty left about the parameters θ_i from an estimate $T(E)$ derived from a single observation.

We computed the gradient for European call options for various strikes and maturities via the FRFT described in the previous section. We implemented the FRFT in HASKELL for all our simulations. The choice of parameter values is consistent with the S&P 500 whose parameter set was estimated in [2] with $\kappa = 5.07$, $\theta = 0.0457$, $\rho = -0.767$, $\sigma = 0.48$, even though our study, according to its qualitative character, holds true for any reasonable choice of parameters. Furthermore, the values of the time dependent parameters r, q, v and x are in accordance with the data for the S&P 500 on March 3, 2014. The continuously-compounded zero-coupon interest rate is $r = 0.167\%$ with dividend yield $q = 1.894\%$. We assumed the S&P 500 is quoted with 1845.73 and a volatility $v = 0.0108$ which was fitted on call prices. All data is obtained from the OPTION METRICS database. Furthermore, the parameters of the FRFT were fixed as $N = 2^{11}$, $\eta = 0.4$, and $\lambda = 3.6549\text{e} - 4$ which yields strike increments of approximately 0.68 within in a range of 1269.5 till 2683.5. The damping factor is $\alpha = 1.5$ throughout the paper.

Throughout this section, we do not consider the variance \hat{v} of the Gaussian Noise. Nevertheless, the figures allow for a comparison of the relative uncertainty of parameter estimates as \hat{v} just corresponds to a global scaling of the Fisher information. Comparably no information from observing a single call price is gained about the relaxation parameter κ , the leverage parameter ρ , and the variance of the variance σ . The estimates of the volatility σ_0 and the long-term mean $\sqrt{\theta}$ of the volatility are about 100 times

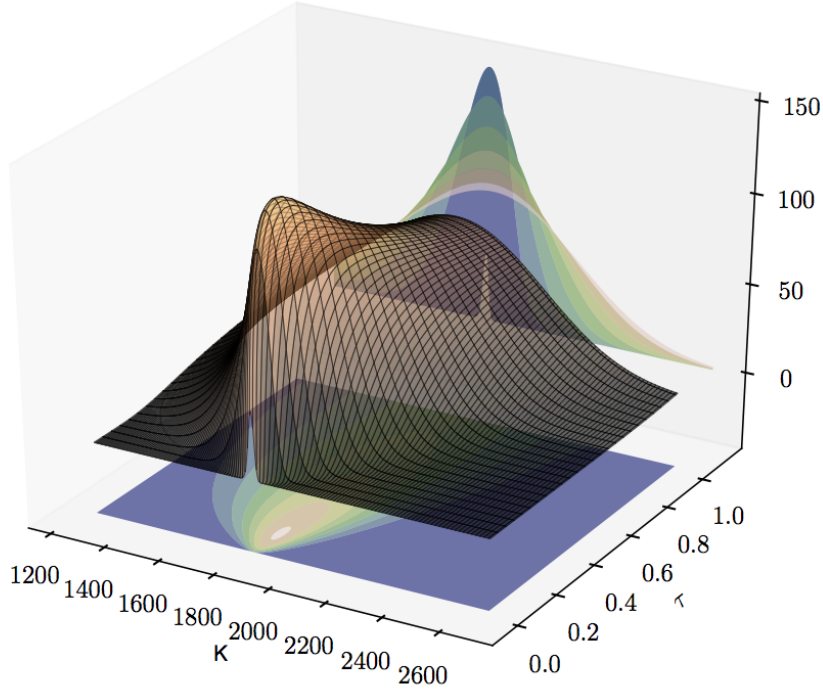


Figure 1: The derivative of Heston's call Price w.r.t. the volatility σ_0

more precise. Interestingly, estimates of σ_0 from call data have an optimal time scale: Vega, the derivative of the call price w.r.t. the volatility attains a global maximum for at the money call options with approximately one month maturity. In general estimates for all parameters are best for at the money call options and uncertainty increases with the distance of the strike from the quoted price of the underlying asset. Due to put-call Parity all these results hold true for put options as well. Hence, in the interest of space their detailed exposure is skipped.

4.2 Vega-Drop

In general parameters are not estimated from a single option price but for various options over an entire time-period. An estimate over a time-period \mathcal{T} changes the picture in a twofold way. First, not only maturity and strikes vary but also volatility σ_t and the price of the underlying asset S_t . Second, since volatility σ_t needs to be estimated for every day $t \in \mathcal{T}$ the parameter vector Θ is no longer $(\sigma_0, \kappa, \sqrt{\theta}, \sigma, \rho)$ but

$$(\sigma_t)_{t \in \mathcal{T}} \oplus (\kappa, \sqrt{\theta}, \sigma, \rho) \quad (4.1)$$

where \oplus denotes concatenation. Consequently, the Fisher information matrix $J(\Theta)$ is an $(m + 4) \times (m + 4)$ matrix where $m = |\mathcal{T}|$ is the number of

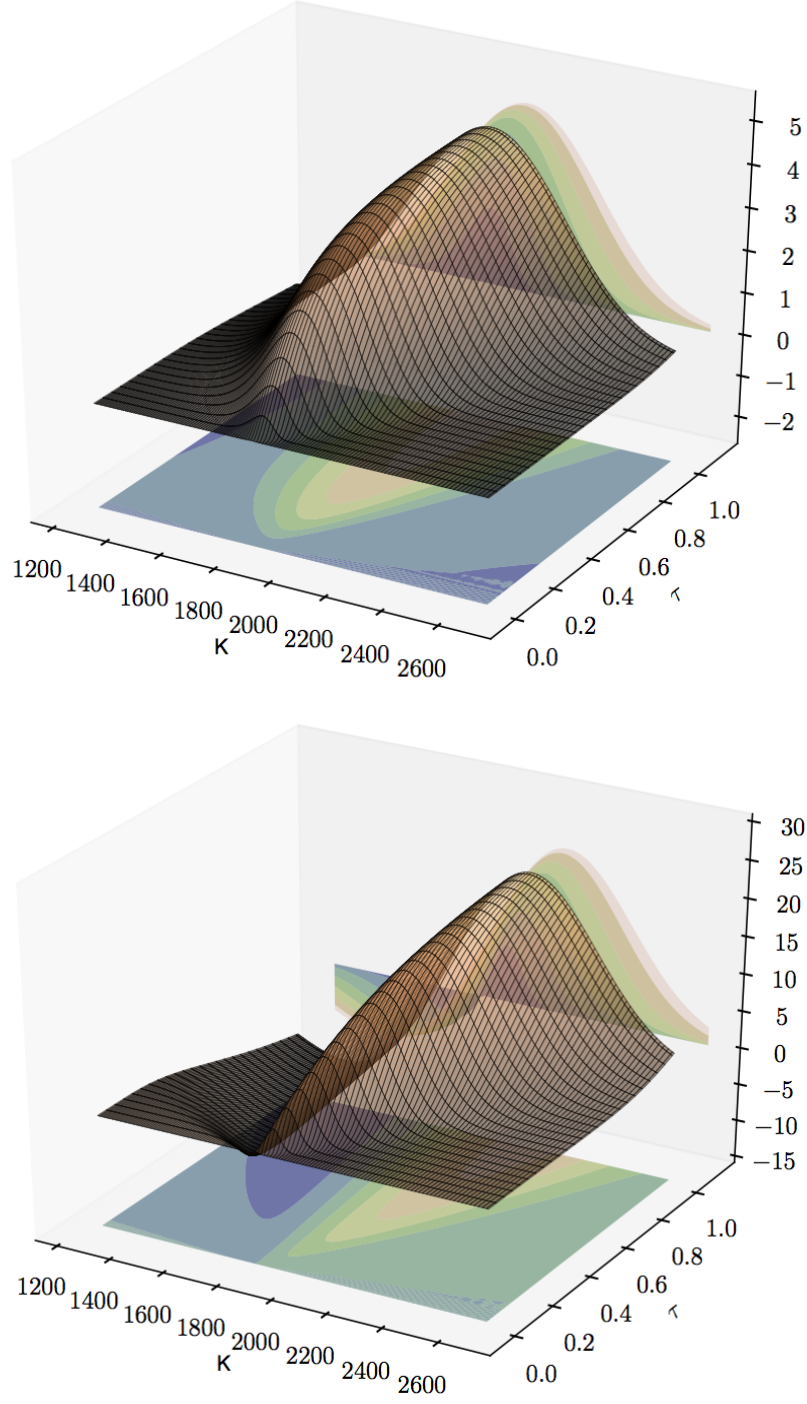


Figure 2: The derivatives of Heston's call Price w.r.t. the relaxation parameter κ , upper figure, and the leverage parameter ρ .

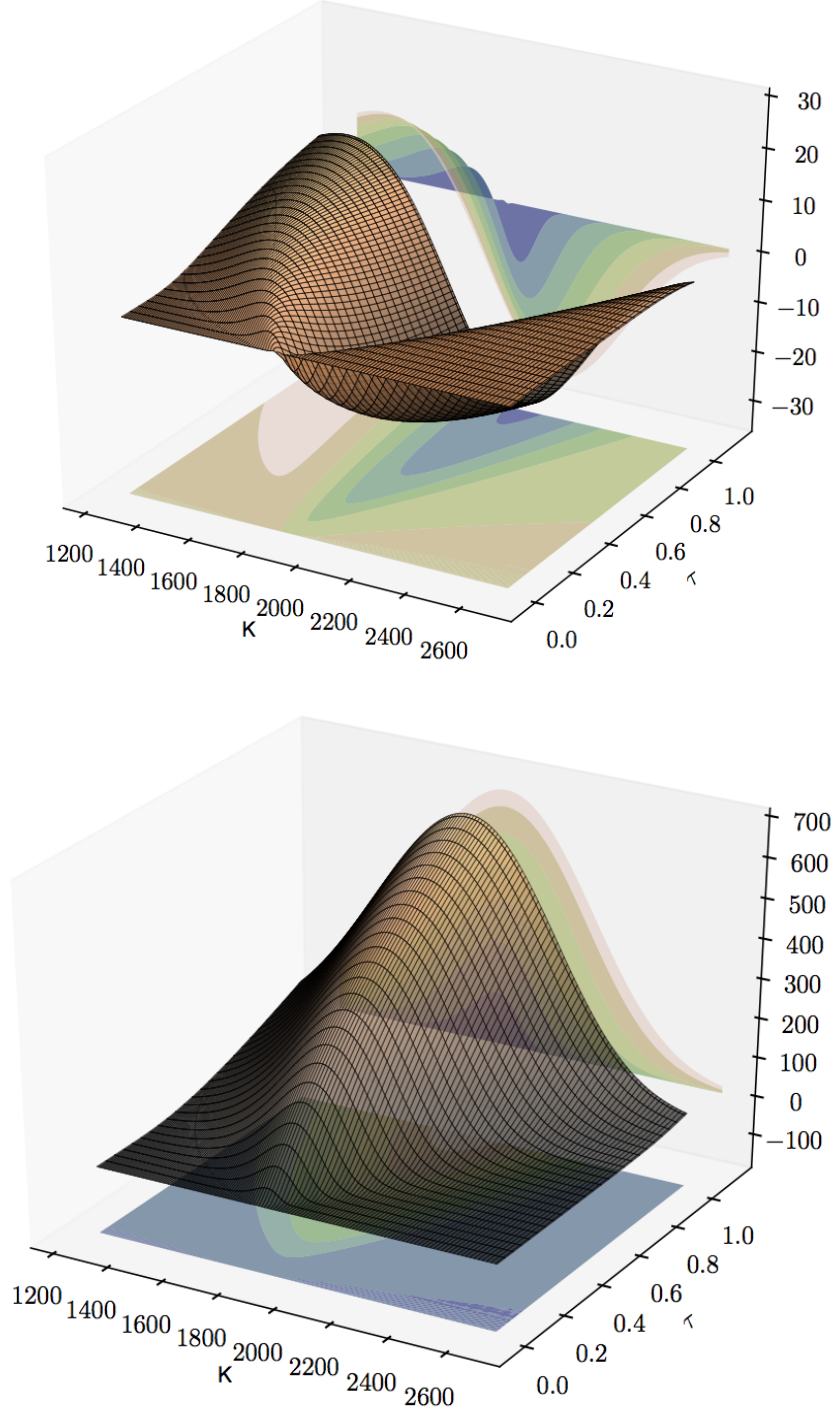


Figure 3: The derivatives of Heston's call Price w.r.t. the volatility of the variance process σ and the long-term mean of the volatility $\sqrt{\theta}$.

days we consider. The first m diagonal elements of $J(\Theta)^{-1}$ provide lower error bounds on the estimates of volatilities $(\sigma_t)_{\mathcal{T}}$. According to lemma 4.1 these diagonal elements are lower bounded by

$$\hat{v} \left(\frac{\partial}{\partial \sigma_t} E_t \right)^{-2} \quad \text{for } t \in \mathcal{T} \quad (4.2)$$

where E_t is an observed option price at time t . Hence, Vega, i.e., the first-order derivative of the option price w.r.t. volatility, plays a crucial role for error estimates for the majority of the parameters in Eq. (4.1).

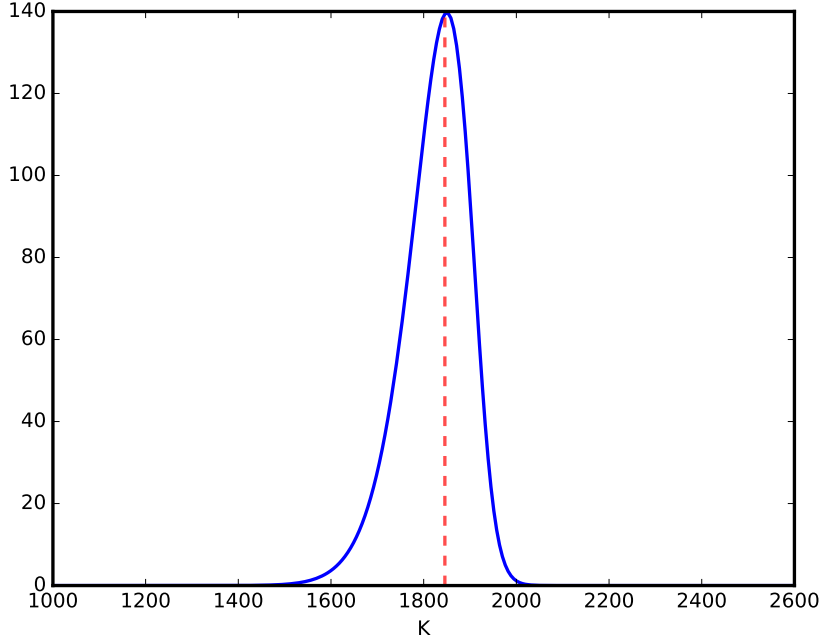


Figure 4: Vega for a European option priced by Heston's model over various strikes. It is strictly positive and has a global maximum near the spot price of the underlying asset (indicated by the red dashed line).

We plotted Vega of a European option traded on the S&P 500 in Fig. (4) over the same parameter set we have already applied for the explicit computation of the Greek-surfaces of the previous subsection: spot price $S = 1845.73$, $r = 0.167\%$, $q = 1.894\%$, $\kappa = 5.07$, $\theta = 0.0457$, $\sigma = 0.48$, $\rho = -0.767$, variance $v = 0.0108$, and maturity $\tau = 30$ days. As in the case of the classical Black Scholes Model, Vega is always positive, i.e., the value of an option increases with volatility, and Vega attains a maximum near the spot price of the underlying. Of greater importance is the dependency of Vega on the variance v if the strike is fixed.

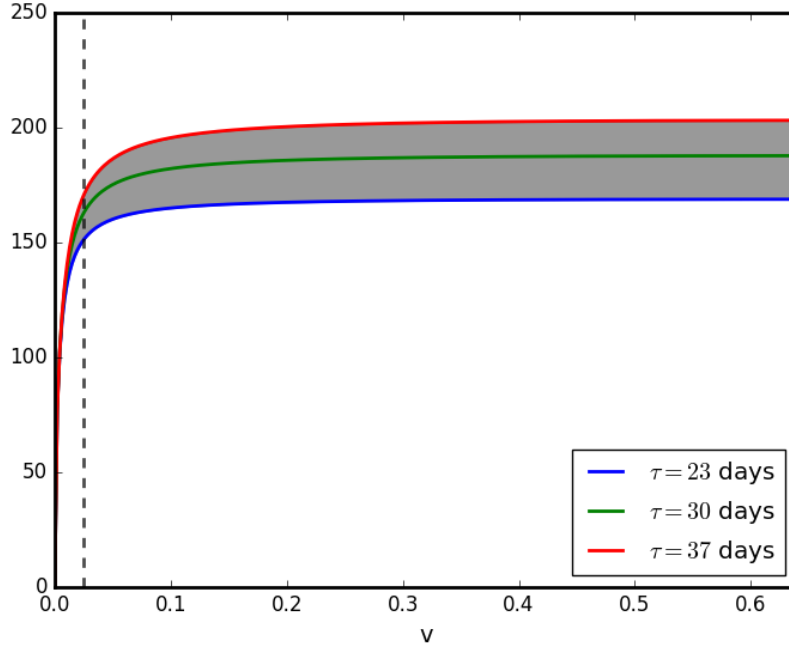


Figure 5: Vega for a European option price by Heston's model over various variances for a fixed strike $K = 1845.73$ at the spot price of the underlying. Vega drops for small variances. The dashed line is at $v = 0.025$.

Fig. (5) plots Vega in Heston's model for an at-the-money option over various variances for three different maturities: 23, 30, and 37 days; the range of maturities of the components of the VIX which is studied in the subsequent section. The curves for different maturities $\tau \in [23, 37]$ are located in the shaded region. One observes a dramatic drop of Vega for small variances v which is indicated by the dashed line located at $v = 0.025$. According to Eq. (4.2) this fact sheds light on variance, and volatility estimates, respectively, from European options: if Vega decreases dramatically the error in estimating the volatility from options becomes large. In the subsequent section we show that this observed Vega-Drop is not only of theoretical but also practical interest if we investigate the quality of the VIX as a volatility proxy.

5 Fisher Information and the VIX

The S&P 500 is a major stock index which is calculated using the prices of approximately 500 component stocks of the biggest companies in the US. The S&P 500, like other indices, employs rules that govern the selection

of the component securities and a formula to calculate index values. The VIX index measures 30-day expected volatility of the S&P 500 index and is comprised by options rather than stocks, with the price of each option reflecting the market's expectation of future volatility. Like conventional indices, the VIX calculation employs rules for selecting component options and a formula to calculate index values. Roughly, the selection filters near- and next-term put and call options with more than 23 days and less than 37 days to expiration and non-vanishing bid – see [1] for further details. With the aid of the OPTION METRICS database we are able to replicate the part of the portfolio¹ of options entering the calculation of the VIX from October 16th 2003 until August 31th 2015. Assuming that all prices of the VIX components are centred on the theoretical price Eq. (2.3) in Heston's model disturbed by additive Gaussian Noise we address two questions to the data: first, how reliably can volatility be inferred from the VIX components; second, does the VIX really measure 30-day expected volatility if one considers the VIX index as a 30-day variance swap traded on the S&P 500? We tackle both questions in terms of Fisher information.

5.1 Inferring volatility

We denote with $\mathcal{E}_t = \mathcal{C}_t \cup \mathcal{P}_t$ the set of all components of the VIX at day t where \mathcal{C}_t denotes the set of call options and \mathcal{P}_t the set of put options, respectively. We introduce the set $\mathcal{T} = \{16/10/2003, 17/10/2003, \dots, 31/08/2015\}$ of all trading days from October 16th 2003 until August 31th 2015. We assume that the price e_t of an option e in the set \mathcal{E}_t is normally distributed

$$p(e_t | \epsilon, x_t, v_t, r_t, \kappa, \theta, \rho, \sigma, k_e, \tau) = \frac{1}{\sqrt{2\pi\hat{v}}} \exp\left(\frac{(e_t - E(\epsilon, x_t, v_t, r_t, \kappa, \theta, \rho, \sigma, k_e, \tau))}{2\hat{v}}\right)$$

with a fixed variance \hat{v} and mean $E(\epsilon, x_t, v_t, r_t, \kappa, \theta, \rho, \sigma, k_e, \tau)$ provided by Eq. (2.2) where x_t is the log-closing-price of the S&P 500 at day t , r_t the zero Coupon Bond yield at day t with maturity τ , and k_e the log-strike of the option. Beside the option prices, also the daily zero Coupon Bond yield r and the closing prices of the S&P 500 are provided by the OPTION METRICS database. Furthermore we adopt the estimate in [2] for the parameters of the stochastic volatility process Eq. (2.1).

$$\kappa = 5.07, \theta = 0.0457, \sigma = 0.48, \rho = -0.767 \quad (5.1)$$

We fit the volatility with daily time-resolution on call data via a maximum likelihood estimate on the call option prices. That is, for every day $t \in \mathcal{T}$,

¹The new VIX is computed based on S&P 500 options which strike dates are renewed in monthly intervals. In addition, weekly options, traded under a different ticker, are used to cover the remaining weeks. These later options are not available in our data base.

the variance v_t , and therefore volatility $\sigma_t = \sqrt{v_t}$, is determined as the maximum of the joint, negative log-likelihood function

$$- \sum_{c \in \mathcal{C}_t} \log(p(c_t | x_t, v_t, r_t, \kappa, \theta, \rho, \sigma, k_c, \tau))$$

which is equivalent to minimize the square-error

$$\sum_{c \in \mathcal{C}_t} \|c_t - C(x_t, v_t, r_t, \kappa, \theta, \rho, \sigma, k_c, \tau)\|^2.$$

This procedure yields a time series $(v_t)_{t \in \mathcal{T}}$ for the variance. Besides, the variance \hat{v} is obtained from maximizing the negative, joint log-likelihood function

$$\sum_{t \in \mathcal{T}} \sum_{c \in \mathcal{C}_t} \log(p(c_t | x_t, v_t, r_t, \kappa, \theta, \rho, \sigma, k_c, \tau))$$

which yields the expression

$$\hat{v} = \frac{1}{N} \sum_{t \in \mathcal{T}} \sum_{c \in \mathcal{C}_t} \|c_t - C(x_t, v_t, r_t, \kappa, \theta, \rho, \sigma, k_c, \tau)\|^2$$

where N is the cardinality of the union $\bigcup_{t \in \mathcal{T}} \mathcal{C}_t$. For our data set of call options from October 16th 2003 until August 31th 2015 we obtain the value

$$\hat{v} = 0.2952.$$

We have assembled all necessary ingredients to tackle the following thought problem: how much information is present in the VIX components about the unknown volatility process $(\sigma_t)_{t \in \mathcal{T}}$, with $\sigma_t = \sqrt{v_t}$, and the parameters $\kappa, \sqrt{\theta}, \sigma$ and ρ ? More precisely, Suppose the parameter vector

$$\Theta = (\sigma_t)_{t \in \mathcal{T}} \oplus (\kappa, \sqrt{\theta}, \sigma, \rho) \tag{5.2}$$

and the joint log-likelihood function

$$\sum_{t \in \mathcal{T}} \sum_{e \in \mathcal{E}_t} \log(p(e_t | \epsilon, x_t, v_t, r_t, \kappa, \theta, \rho, \sigma, k_e, \tau)).$$

What is the lower bound on the loss function Eq. (3.1) of an estimator of Θ provided all VIX components $\bigcup_{t \in \mathcal{T}} \mathcal{E}_t$ from October 16th 2003 until August 31th 2015 as data? According to the Cramér-Rao inequality the inverse of the Fisher information matrix $J(\Theta)$ yields the answer. $J(\Theta)$ adopts in the present context the block structure

$$J(\Theta) = \frac{1}{\hat{v}} \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}. \tag{5.3}$$

$\mathbf{A}_{11} = (a_{ij})$ is an $m \times m$ diagonal matrix, where m is the cardinality of \mathcal{T} s.t.

$$a_{ii} = (\partial_{\sigma_{t_i}} E_i)^2 = 4v_{t_i} (\partial_{v_{t_i}} E_i)^2$$

where we write

$$E_i = \sum_{e \in \mathcal{E}_{t_i}} E(\epsilon, x_{t_i}, v_{t_i}, r_{t_i}, \kappa, \theta, \rho, \sigma, k_e, \tau). \quad (5.4)$$

\mathbf{A}_{12} is an $m \times 4$ matrix,

$$\mathbf{A}_{12} = \begin{pmatrix} \partial_{\sigma_{t_0}} E_0 \partial_{\kappa} E_0 & \partial_{\sigma_{t_0}} E_0 \partial_{\sqrt{\theta}} E_0 & \partial_{\sigma_{t_0}} E_0 \partial_{\sigma} E_0 & \partial_{\sigma_{t_0}} E_0 \partial_{\rho} E_0 \\ \vdots & \vdots & \vdots & \vdots \\ \partial_{\sigma_{t_m}} E_m \partial_{\kappa} E_m & \partial_{\sigma_{t_m}} E_m \partial_{\sqrt{\theta}} E_m & \partial_{\sigma_{t_m}} E_m \partial_{\sigma} E_m & \partial_{\sigma_{t_m}} E_m \partial_{\rho} E_m \end{pmatrix},$$

and $\mathbf{A}_{21} = \mathbf{A}_{12}^T$. Finally, \mathbf{A}_{22} is the 4×4 matrix

$$\mathbf{A}_{22} = \sum_{i=1}^m \nabla E_i \nabla E_i^T$$

where $\nabla E_i = (\partial_{\kappa} E_i, \partial_{\theta} E_i, \partial_{\sigma} E_i, \partial_{\rho} E_i)^T$ denotes the column gradient vector of E_i w.r.t. the parameters of the stochastic process Eq. (2.1). If we assume that the previously fitted variance time-series $(v_t)_{t \in \mathcal{T}}$ and the parameter set Eq. (5.1) provide an estimator of Θ , we can evaluate all the derivatives for computing the Fisher information matrix $J(\Theta)$ Eq. (5.3) whose inverse yields a lower bound on the loss function of this estimator. Fig. (6) presents the volatility time-series $(\sigma_t)_{t \in \mathcal{T}}$ obtained from the previously described fit on components entering the calculation of the VIX along with the uncertainty left.

Since the uncertainty left is negligible the shaded tube $([\sigma_t - \beta_t, \sigma_t + \beta_t])_{t \in \mathcal{T}}$ (95% credibility region) mantling the plot of the volatility time-series $(\sigma_t)_{t \in \mathcal{T}}$ is only visible in a close-up presented in Fig. (6) as well. The close-up shows volatility from February 23rd 2010 until March 10th 2010 within a range from 0.135 till 0.180. The boundaries $(\beta_t)_{t \in \mathcal{T}}$ of the credibility region are obtained from the first m entries of the diagonal of the inverse $J(\Theta)^{-1} = (J(\Theta)_{ij}^{-1})$ of the Fisher information matrix (Recall, m is the number of trading days we consider, i.e. the cardinality of the set \mathcal{T}):

$$\beta_{t_i} = 2\sqrt{J(\Theta)_{ii}^{-1}} \quad \text{for all } i = 1, \dots, m.$$

That is, in terms of the Cramér-Rao inequality, β_t represents a lower bound on the double standard deviation of the volatility estimate σ_t . On average the double standard deviation is

$$\bar{\beta} = \frac{1}{m} \sum_{i=0}^m \beta_{t_i} = 0.0014.$$

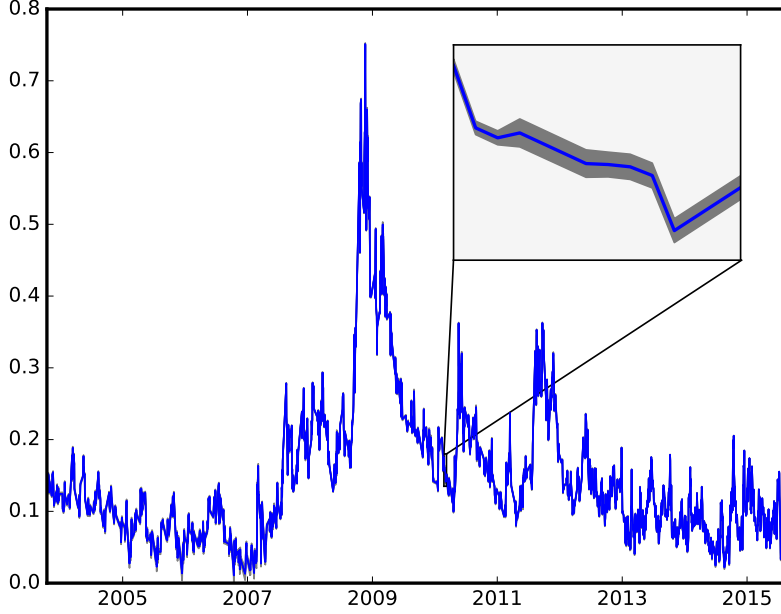


Figure 6: Volatility of the S&P 500. The close-up presents the volatility estimate along with its double standard deviation obtained from Fisher information matrix Eq. (5.3).

The last four entries of the diagonal of $J(\Theta)^{-1} = (J(\Theta)_{ij}^{-1})$ yield a lower bound on the variance of the estimates of the parameters $\kappa, \sqrt{\theta}, \sigma$ and ρ , respectively. Along with the parameter estimates of [2] we obtain table 1. Fitting the parameters $\kappa, \sqrt{\theta}, \sigma$ and ρ on options over a sufficiently long time-window yields fairly accurate estimates of them as well.

Overall, estimates of hidden volatility and the parameters determining the stochastic process Eq. (2.1) from option data appear reliable. Doubts are shed on these results if relative errors are considered instead of absolute ones. Fig. (7) presents the relative uncertainty left, that is, the time-series $(\beta_t/\sigma_t)_{t \in \mathcal{T}}$. Apparently, the uncertainty becomes overwhelming, if the volatility drops. This is in accordance with the discussion in the previous section where we observed that Vega drops if volatility falls below a critical value. From the block structure Eq. (5.3) of the Fisher information matrix and the proof of lemma 4.1 follows

$$J(\Theta)_{ii}^{-1} \geq \frac{\hat{v}}{(\partial_{\sigma_{t_i}} E_i)^2}$$

	Estimate	Standard Error
κ	5.07	$4.0e - 2$
$\sqrt{\theta}$	0.214	$6.5e - 4$
σ	0.48	$9.4e - 4$
ρ	-0.767	$7.5e - 4$

Table 1: The parameters of Heston’s model along with their standard errors obtained from the Fisher information matrix if we assume they were estimated from the components of the VIX between October 16th 2003 until August 31th 2015.

with E_i provided by Eq. (5.4), and therefore

$$\beta_{t_i} \geq \frac{2\sqrt{\hat{v}}}{\partial_{\sigma_{t_i}} E_i}$$

for all $i = 1, \dots, m$. Hence, there is a lower bound of the error of the volatility σ_t proportional to the inverse of a sum of Vegas. If this sum shrinks the error becomes large.

5.2 The VIX as variance swap

The VIX is quoted in percentage points and translates, roughly, to the expected movement (with the assumption of a 68% likelihood, i.e., one standard deviation) in the S&P 500 index over the next 30-day period, which is then annualized. According to Carr and Wu [11] the VIX can also be considered as 30 day variance swap on the S&P 500. Recall Eq. (2.1) that the variance of the Heston model is driven by the CIR process

$$dv_t = \kappa(\theta - v_t)dt + \sigma\sqrt{v_t}dW_t$$

and consequently, that the expected value of v_t conditional on v_s ($s < t$) is

$$\mathbb{E}[v_t|v_s] = v_s e^{-\kappa(t-s)} + \theta \left(1 - e^{-\kappa(t-s)}\right) = (v_s - \theta)e^{-\kappa(t-s)} + \theta.$$

In the sequel, we make use of $\mathbb{E}[v_t|v_s]$ but with $s = 0$. It is useful to denote this quantity as \hat{v}_t

$$\hat{v}_t = \mathbb{E}[v_t|v_0] = (v_0 - \theta)e^{-\kappa t} + \theta.$$

It is also useful to define the total (integrated) variance \hat{w}_t as

$$\hat{w}_t = \int_0^t v_s ds = (v_0 - \theta) \frac{1 - e^{-\kappa t}}{\kappa} + \theta t.$$

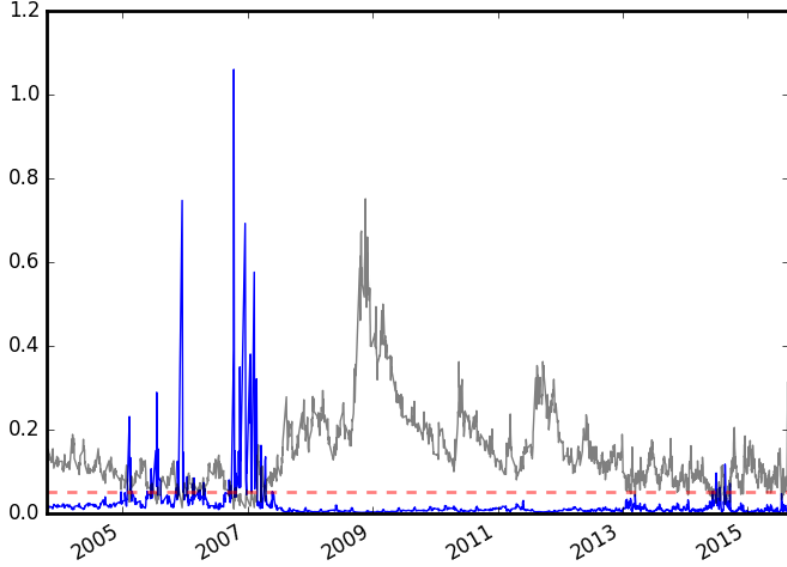


Figure 7: The relative error of the volatility estimate (blue line) and volatility itself (transparent black line). If volatility drops below a certain level (dashed, transparent red line), option data yield no information about volatility.

As explained by Gatheral [19], a variance swap requires an estimate of the future variance over the $(0, T)$ time period. This can be obtained as the conditional expectation of the integrated variance. A fair estimate of the total variance is therefore

$$\begin{aligned} \mathbb{E} \left[\int_0^T v_t dt | v_0 \right] &= \int_0^T \mathbb{E}[v_t | v_0] dt \\ &= \int_0^T (v_0 - \theta) e^{-\kappa t} + \theta dt = (v_0 - \theta) \frac{1 - e^{-\kappa T}}{\kappa} + \theta T \end{aligned}$$

which is simply \hat{w}_T . Since this represents the total variance over $(0, T)$, it must be scaled by T in order to represent a fair estimate of annual variance (assuming that T is expressed in years). Hence, the strike variance K_{var}^2 for a variance swap is obtained by dividing this last expression by T

$$K_{\text{var}}^2 = (v_0 - \theta) \frac{1 - e^{-\kappa T}}{\kappa T} + \theta.$$

Returning to Carr's and Wu's interpretation of the VIX [11] the VIX-time series $(\text{VIX}_t)_{t \in \mathcal{T}}$ is the time-series

$$K_{\text{var},t} = \sqrt{(v_t - \theta) \frac{1 - e^{-\kappa T}}{\kappa T}} + \theta \quad \text{for } t \in \mathcal{T}$$

where $(v_t)_{t \in \mathcal{T}}$ denotes the variance of the S&P 500 at day t and $T = 30/365$. Thus, Fisher information provides an uncertainty estimate on the VIX, considered as 30 days variance swap, estimated from its components, i.e., near- and next-term put and call options with more than 23 days and less than 37 days to expiration and non-vanishing bid. We only have to transform the Fisher information matrix already computed in the previous subsection according to the rules Eq. (3.2). In terms of Eq. (3.2) in section 2.1 we have

$$\begin{aligned} \mathbf{\Lambda} &= (K_{\text{var},t})_{t \in \mathcal{T}} \oplus (\kappa, \sqrt{\theta}, \sigma, \rho) \\ \mathbf{\Theta} &= (\sigma_t)_{t \in \mathcal{T}} \oplus (\kappa, \sqrt{\theta}, \sigma, \rho) \end{aligned}$$

and

$$J(\mathbf{\Lambda}) = D^T J(\mathbf{\Theta}) D.$$

Hence, the transformation matrix D in Eq. (3.2) adopts the block form

$$D = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}.$$

\mathbf{A}_{11} is an $m \times m$ diagonal matrix with entries

$$\begin{aligned} a_{ii} &= \frac{\partial \sigma_{t_i}}{\partial K_{\text{var},t_i}} = \frac{\partial \sigma_{t_i}}{\partial v_{t_i}} \frac{\partial v_{t_i}}{\partial K_{\text{var},t_i}} \\ &= \frac{1}{2\sigma_{t_i}} \frac{\partial}{\partial K_{\text{var},t_i}} \left(\frac{(K_{\text{var},t_i}^2 - \theta)\kappa T}{1 - e^{-\kappa T}} + \theta \right) = \frac{1}{\sigma_{t_i}} \frac{K_{\text{var},t_i} \kappa T}{1 - e^{-\kappa T}} \end{aligned}$$

for $i = 1, \dots, m$ where m is the number of trading days $t_i \in \mathcal{T}$ from October 16th 2003 until August 31th 2015. \mathbf{A}_{12} is an $m \times 4$ -matrix with entries

$$\begin{pmatrix} \frac{\partial \sigma_{t_0}}{\partial \kappa} & \frac{\partial \sigma_{t_0}}{\partial \sqrt{\theta}} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial \sigma_{t_m}}{\partial \kappa} & \frac{\partial \sigma_{t_m}}{\partial \sqrt{\theta}} & 0 & 0 \end{pmatrix}$$

where

$$\begin{aligned} \frac{\partial \sigma_{t_i}}{\partial \kappa} &= \frac{\partial \sigma_{t_i}}{\partial v_{t_i}} \frac{\partial v_{t_i}}{\partial \kappa} = \frac{(K_{\text{var},t_i}^2 - \theta)T}{2\sigma_{t_i}} \frac{\partial}{\partial \kappa} \left(\frac{\kappa}{1 - e^{-\kappa T}} \right) \\ &= \frac{(K_{\text{var},t_i}^2 - \theta)T}{2\sigma_{t_i}} \frac{1 - (1 + \kappa T)e^{-\kappa T}}{(1 - e^{-\kappa T})^2} \\ \frac{\partial \sigma_{t_i}}{\partial \sqrt{\theta}} &= \frac{\partial \sigma_{t_i}}{\partial v_{t_i}} \frac{\partial \theta}{\partial \sqrt{\theta}} \frac{\partial v_{t_i}}{\partial \theta} = \frac{\sqrt{\theta}}{\sigma_{t_i}} \left(1 - \frac{\kappa T}{1 - e^{-\kappa T}} \right). \end{aligned}$$

Finally, $\mathbf{A}_{21} = \mathbf{A}_{12}^T$ and \mathbf{A}_{22} is the 4×4 identity matrix. Similar to figure Fig. (6) we plot Fig. (8) the time-series $(K_{\text{var},t})_{t \in \mathcal{T}}$ along with its 95% credibility region $[K_{\text{var},t} - \beta_t, K_{\text{var},t} + \beta_t]$ where

$$\beta_{t_i} = 2\sqrt{J(\mathbf{\Lambda})_{ii}^{-1}} \quad \text{for } i = 1, \dots, m.$$

Furthermore, the historical VIX is plotted. The value $K_{\text{var},t}$ is systematically smaller than the realized VIX.

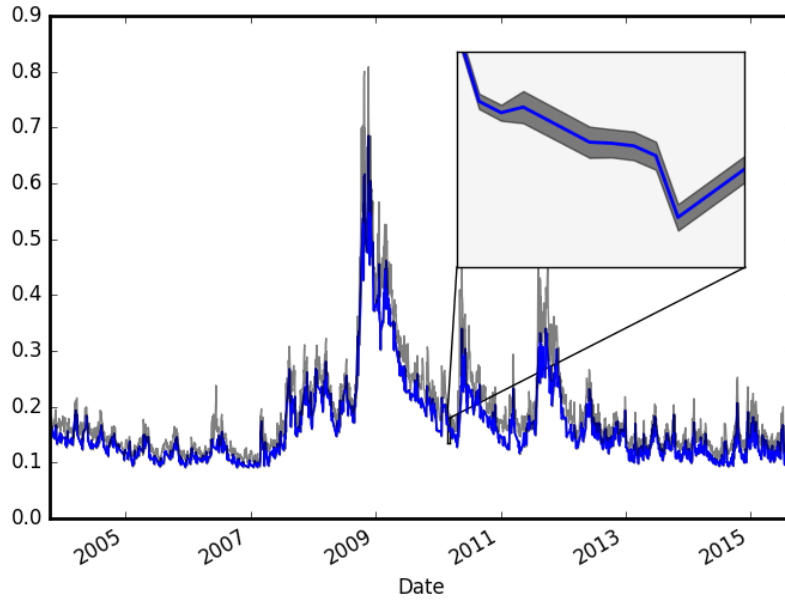


Figure 8: The time series $(K_{\text{var},t})_{t \in \mathcal{T}}$. The close-up shows the graph along with its 95% credibility region. The Black transparent line shows the historical VIX.

As for the volatility the error is negligible and on average we have

$$\bar{\beta} = \frac{1}{m} \sum_{i=0}^m \beta_{t_i} = 8.9e - 04.$$

Compared to the volatility estimates from options there is big difference if we consider the relative error, that is, the time-series $\beta_t/K_{\text{var},t}$ for $t \in \mathcal{T}$ in Fig. (9). The relative error never exceeds 3%.

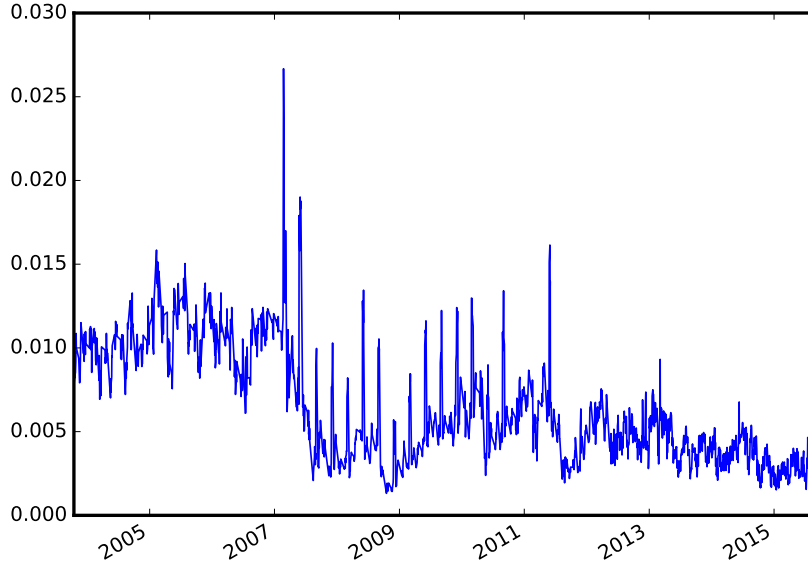


Figure 9: The time series $(\beta_t / K_{\text{var},t})_{t \in \mathcal{T}}$.

6 Conclusion

Here, we have addressed the question of how reliably volatility can be estimated from option price data. To this end, we computed the Fisher information matrix of Heston's stochastic volatility model. Thanks to the analytic tractability of Heston's model, the Fisher information can be expressed in Fourier integrals giving the Heston Greeks.

Our investigations lead to the following insights: First, options at the money are most informative about volatility while almost no information can be obtained from options that are far out of the money. Second, low volatilities are hard to estimate as Vega almost vanishes in this case, making it impossible to extract information from option prices. Third, volatility estimation from market data as exemplified on S&P 500 index options is reliable most of the time, with occasional large relative errors for very low volatilities. We might speculate that this could lead to overconfident estimates of portfolio risk especially in times of calm financial markets. Nevertheless, the VIX index itself, reflecting the average volatility over the next month, proves to be an accurate assessment of volatility.

Overall, this work complements our previous findings [31] regarding the low information content of stock returns. There, we showed that in general

at least secondly quoted return data is necessary to infer volatility successfully. We guessed that option price data could lead to much more reliable volatility estimates as confirmed by the analysis presented here.

7 Funding

Nils Bertschinger and Oliver Pfante thank Dr. h.c. Maucher for funding their positions.

References

- [1] The cboe volatility index - vix.
- [2] Y. Aït-Sahalia and R. Kimmel. Maximum likelihood estimation of stochastic volatility models. *Journal of Financial Economics*, 83(2):413–452, 2007.
- [3] H. Albrecher, P. Mayer, W. Schoutens, and J. Tistaert. The little heston trap. *Wilmott Magazine*, pages 83–92, 2007.
- [4] F. Black and M. Scholes. The Pricing of Options and Corporate Liabilities. *Journal of Political Economy*, 81(3):637, 1973.
- [5] T. Bollerslev, J. Litvinova, and G. Tauchen. Leverage and volatility feedback effects in high-frequency data. *Journal of Financial Econometrics*, 4(3):353–384, 2006.
- [6] J. P. Bouchaud, a. Matacz, and M. Potters. Leverage effect in financial markets: the retarded volatility model. *Physical review letters*, 87(22):228701, 2001.
- [7] J.-P. Bouchaud and M. Potters. Theory of Financial Risk and Derivative Pricing. *Theory of Financial Risk and Derivative Pricing*, page 379, 2003.
- [8] M. Brenner and D. Galai. New financial instruments for hedging changes in volatility. *Financial Analysts Journal*, 45(4):61–65, 1989.
- [9] M. Brenner and D. Galai. Hedging Volatility in Foreign Currencies. *The Journal of Derivatives*, 1(1):53–59, 1993.
- [10] P. Carr and D. Madan. Option valuation using the fast fourier transform. *Journal of Computational Finance*, 2:61–73, 1999.
- [11] P. Carr and L. Wu. A tale of two indices. *NYU Working Paper No. SC-CFE-04-01*, 2014.

- [12] K. Chourdakis. Option pricing using the fractional fft. *Journal of Computational Finance*, 8(2):1–18, 2005.
- [13] A. A. Christie. The stochastic behavior of common stock variances: Value, leverage and interest rate effects. *Journal of Financial Economics*, 10(4):407–432, 1982.
- [14] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley-Interscience, 2 edition, July 2006.
- [15] Z. Ding, C. W. Granger, and R. F. Engle. A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, 1(1):83–106, 1993.
- [16] A. Dragulescu and V. M. Yakovenko. Probability distribution of returns in the Heston model with stochastic volatility. *Quantitative Finance*, 2:443–453, 2002.
- [17] O. E. Euch and M. Rosenbaum. The characteristic function of rough heston models. *arXiv:1609.02108 [q-fin.MF]*, 2016.
- [18] M. Forde, A. Jacquier, and R. Lee. The small-time smile and term structure of implied volatility under the heston model. *SIAM Journal on Financial Mathematics*, 3(1):690–708, 2012.
- [19] J. Gatheral. *The volatility surface: a practitioner’s guide*. John Wiley & Sons, 2011.
- [20] S. L. Heston. A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options. *The Review of Financial Studies*, 6(2):327–343, 1993.
- [21] A. Jacquier and P. Roome. The small-maturity heston forward smile. *SIAM Journal on Financial Mathematics*, 4(1):831–856, 2013.
- [22] A. Jacquier and P. Roome. Large-maturity regimes of the heston forward smile. *Stochastic Processes and their Applications*, 126(4):1087–1123, 2016.
- [23] A. Janek, T. Kluge, R. Weron, and U. Wystup. *FX smile in the Heston model*, pages 133–162. Springer, statistical tools for finance and insurance edition, 2011.
- [24] B. LeBaron. Stochastic Volatility as a Simple Generator of apparent financial power laws and long memory. *Quantitative Finance*, 6(1):627–631, 2001.
- [25] A. W. Lo. Long Term Memory in Stock Market Prices. *Econometrica*, 59(5):1279–1313, 1991.

- [26] B. B. Mandelbrot. *The variation of certain speculative prices*. Springer, 1997.
- [27] A. Mazzon and A. Pascucci. The forward smile in local-stochastic volatility models. *Available at SSRN 2560300*, 2015.
- [28] J. F. Muzy, J. Delour, and E. Bacry. Modelling fluctuations of financial time series : from cascade process to stochastic volatility model. *Eur. Phys. J. B*, 17:537–548, 2000.
- [29] J. Perello, R. Sircar, and J. Masoliver. Option pricing under stochastic volatility: the exponential Ornstein-Uhlenbeck model. page 26, 2008.
- [30] K. B. Petersen and M. S. Pedersen. The matrix cookbook. <http://matrixcookbook.com>, 2012.
- [31] O. Pfante and N. Bertschinger. Volatility inference and return dependencies in stochastic volatility models. *arXiv:1610.00312 [q-fin.MF]*, 2016.
- [32] S.-H. Poon. The heston option pricing model. *Unpublished Draft*, 2009.
- [33] F. D. Rouah. *The Heston Model and Its Extensions in Matlab and C#*. Wiley, 2013.